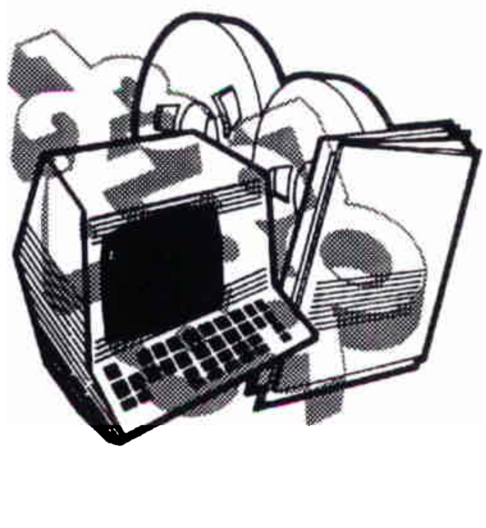
Bureau of the Census

1995 Annual Research Conference

March 19–21, 1995

Key Bridge Marriott Hotel
1401 Lee Highway
Arlington (Rosslyn), Virginia 22209

Proceedings



U.S. Department of Commerce
Economics and Statistics Administration
BUREAU OF THE CENSUS

A SIMULATION STUDY TO EVALUATE THE PERFORMANCE OF MODEL-BASED MULTIPLE IMPUTATIONS IN NCHS HEALTH EXAMINATION SURVEYS

T. Ezzati-Rice[1], W. Johnson[1], M. Khare[1], R. Little[2], D. Rubin[3] and J. Schafer[4]
[1]National Center for Health Statistics, [2]University of Michigan, [3]Harvard University,
[4]Pennsylvania State University

ABSTRACT

The Third National Health and Nutrition Examination Survey (NHANES-III) is the third in a series of periodic surveys conducted by the National Center for Health Statistics to assess the health and nutritional status of the U.S. population. This survey is subject to non-negligible levels of unit and item nonresponse, both in its interview and its examination components. In previous surveys, nonresponse was handled primarily using weighting adjustments. Increasing levels of nonresponse in NHANES, and inconsistencies in analyses of NHANES data attributable to differing treatments of the missing values, led to the desire to develop imputation methods for NHANES-III and subsequent NHANES surveys.

Previous research developed and compared alternative missing-data adjustments based on single and multiple imputation of missing data on NHANES-III data, and assessed mixed weighting and imputation strategies. In Little and Rubin (1992), it was proposed that nonresponse for individuals missing both the questionnaire and MEC examination be handled by the first-stage weight adjustments developed in Ezzati and Khare (1992). Important variables in the data set subject to other forms of nonresponse would be multiply-imputed (Rubin 1987) using a method based on the general location model (Schafer 1991), with 5 multiple imputes for each missing value (MIGLOM, for Multiple Imputation under the General Location Model). Missing-values of less important variables in the analysis file would be replaced by missing-value codes.

Previous work demonstrated the feasibility of this strategy, and provided useful descriptive information for assessing and comparing imputation methods, but it did not provide objective information about the operating characteristics of the model-based imputation method -- for example, do $100(1-\alpha)$ percent confidence intervals created from the multiple-imputed data sets developed using MIGLOM really cover the target quantity in a proportion $1-\alpha$ of repeated samples for NHANES?

This article reports the results of a simulation study designed to assess frequentist properties of MIGLOM. Our simulations were based on a hypothetical population constructed from previous NHANES data sets to resemble populations surveyed by NHANES. Samples were then drawn from the population by a stratified random sampling method that captured some (although not all) of the characteristics of NHANES sampling designs. Missing data were created in ways that mimicked the mechanism and patterns of nonresponse in NHANES-III. Our multiple imputation methods were then applied to the resulting incomplete data, and the validity of multiple imputations assessed by computing the width of intervals and their coverage of underlying population quantities. The results show that MIGLOM is remarkably successful in creating valid design-based repeated-sampling inferences.

1. INTRODUCTION

1.1 The problem. The Third National Health and Nutrition Examination Survey (NHANES-III) is the third in a series of periodic surveys conducted by the National Center for Health Statistics to assess the health and nutritional status of the U.S. population. The NHANES-III survey began in October 1988 and ended in October 1994. Phase 1 of the survey, conducted in 1988-1991, involved data collection on a national probability sample of the civilian non-institutionalized U.S. population, and was the basis for this project. This survey is subject to non-negligible levels of unit and item nonresponse, both in its interview and its examination components. In previous surveys, nonresponse was handled primarily by sample weight adjustments. Increasing levels of nonresponse in NHANES, and inconsistencies in analyses of NHANES data attributable to differing treatments of the missing values, led to the desire to develop imputation methods for NHANES-III and subsequent NHANES surveys.

Variables in NHANES-III can be usefully classified into three groups:

1. Sample frame / household screening variables
2. Interview variables (family and health history variables)
3. Mobile Examination Center (MEC) variables

The sample frame / household screening variables can be treated essentially as fully observed. Missing data in the interview variables are referred to here as interview nonresponse. The interview data consist of family questionnaire variables, and health variables obtained for sampled individuals. For NHANES-III, Little and Rubin (1992) report 15%-16% of values missing on the family questionnaire variables, and 18%-23% of values missing on the selected adult questionnaire variables, the latter reflecting somewhat higher amounts of item nonresponse. Missing data in the MEC variables are referred to here as examination nonresponse. Little and Rubin (1992) report levels of nonresponse of 31%-34% of examination nonresponse in NHANES-III. The MEC examination involves *components* corresponding to related sets of measurements. Item nonresponse for the MEC often arises when all the variables in a particular component are missed. Thus, item nonresponse for the MEC variables is classified as either *component* nonresponse, where an individual is examined but all the variables in one component of the exam are missing, or *item-within-component* nonresponse for particular items, typically a minor problem.

When missing or present as a set, the three blocks of variables (screening, interview, examination) have an approximately monotone structure, with screening variables fully observed, questionnaire variables missing when the interview is not conducted, and examination variables missing when either (a) the interview is not conducted or (b) the interview is conducted but the MEC examination does not take place. However, item nonresponse for interview data, and component and item-within component nonresponse for MEC data, spoil this monotone structure.

1.2 Previous work. Little and Rubin (1992) summarizes research conducted by Datametrics Research in collaboration with NCHS, Joe Schafer and Westat, Inc. References include Ezzati-Rice, Khare, Rubin, Little and Schafer (1993), Khare, Little, Rubin and Schafer (1993), Fahimi and Judkins (1993), and Schafer, Khare and Ezzati-Rice (1993). That work developed and compared alternative missing-data adjustments based on single and multiple imputation of missing data on NHANES-III data, and assessed mixed weighting and imputation strategies. The data set used to develop the imputation models consisted of a subset of the Phase 1 data of NHANES-III. The subset was restricted to adults over 17 years old, and included 6 completely-observed sample frame/household screening variables, 22 interview variables, missing in about one out of every five cases, and 12 MEC variables from three components of the examination, missing in about one out of every three cases .

Three imputation methods were developed as part of that project. Westat, Inc. imputed about 690 missing values for six MEC variables, namely log height, log weight, systolic and diastolic blood pressure, total serum cholesterol and HDL cholesterol, using two closely-related regression imputation methods. Datametrics Research and Joe Schafer multiply imputed 23 variables in the adult questionnaire and MEC, for over 4000 cases involving all types of nonresponse, using a method (MIGLOM) based on Bayesian simulation for the general location model for mixed normal and categorical data (Little and Schluchter 1985; Schafer 1991). The major differences in the number of values imputed reflect in part differences in strategies regarding how much of the missing-data problem should be handled by weighting adjustments and how much by imputation. Specifically, Westat proposed a strategy (ILO -- imputation-LO) that confines imputation to item nonresponse for individuals who received the MEC examination, whereas Datametrics and Schafer applied a strategy (IHI -- imputation-HI) that also imputes all the other sources of unit and item nonresponse.

Comparisons of the imputes from the MIGLOM and Westat methods on the subset of values imputed by Westat indicated that both methods by and large produced reasonable imputations of the missing values, although Schafer's analysis did uncover the need to edit out some implausible respondent values. The MIGLOM method was recommended, in view of its ability to handle general patterns of missing data, and to provide multiple imputations of the missing values that allow the uncertainty due to nonresponse to be incorporated into the analysis using standard complete-data methods.

In Little and Rubin (1992), the relative strengths of weighting and imputation were debated, and a middle course (IMID) was proposed for current implementation. In IMID, nonresponse for individuals missing the questionnaire and MEC examination would be handled by the first-stage weight adjustments developed in Ezzati and Khare (1992). Important variables in the data set subject to other forms of nonresponse would be multiply-imputed using the MIGLOM method, with 5 multiple imputes for each missing value. Missing-values of less important variables in the analysis file would be replaced by missing-value codes. In future, multiple imputation methodology could be expanded to more variables and more incomplete cases, as the method becomes established with users, documentation is refined, and computing limits are reduced by advances in statistical computing.

1.3 Goals of this project. Our previous project demonstrated the feasibility of this IMID strategy, and provided useful descriptive information for assessing and comparing imputation methods, but it did not provide objective information about the operating characteristics of the MIGLOM method -- for example, do $100(1-\alpha)$ percent confidence intervals created from the multiply-imputed data sets really cover the target quantity in a proportion $1-\alpha$ of repeated samples for NHANES? It was recognized that to convince skeptics, what was needed was an honest frequentist evaluation via simulation to assess the methods. The current project reports the results of a simulation study designed to accomplish this objective.

Perhaps the most challenging aspect of studies of this kind is to generate simulation populations that are useful representations of the target populations under study. Our approach had some novel features, including the pooling data from a number of prior HHANES and NHANES surveys, and a disproportionate stratified random sampling scheme that corrected sampling distortions in the pooled population. The sampling method captured some (although not all) the characteristics of NHANES sampling designs. Certain clustering features could not be captured for reasons that will be explained below. Missing data were created in ways that mimicked the mechanism and patterns of nonresponse in NHANES-III. Our multiple imputation methods were then applied to the resulting incomplete data, and the validity of multiple imputations assessed by computing widths of intervals and their coverage of underlying population quantities. The results show clearly that the multiple imputation methods are remarkably successful in creating valid repeated-sampling inferences.

A more comprehensive project report (Little and Rubin 1995) includes detailed memoranda describing the work and simulation results. This paper summarizes the multiple imputation method, the simulation methods and results.

2. CONSTRUCTION OF THE SIMULATION POPULATION AND SAMPLES, AND MISSING DATA.

2.1 General Objective. The goal was to create a hypothetical population that was considerably larger than the intended sample of approximately 12,000 adults for the planned NHANES 97+. It was not intended to reflect accurately distributions of survey outcomes in the real population of the United States at any particular time, but rather to provide a useful basis for assessing alternative statistical procedures to be applied in NHANES settings. That is, if a procedure works well on this population with a known answer, we should be confident that it will also work well on the real population. If it works poorly on this artificial population, then we should not believe in its propriety for the true population.

By combining information from a set of prior NHANES surveys, we obtained a population that is more realistic than a population simulated under a particular stochastic model, because the data for each population unit are the same as the data for a real person and do not involve modeling assumptions. Moreover, this population is also superior to a population bootstrapped from any one sample such as NHANES-III, since such populations have more excessive replication of units and do not adequately reflect variability and diversity in the tails of distributions. In addition to problems with bootstrap samples of statistics related to diversity, bootstrapped summary statistics are too variable in that they include variability due to sampling as well as true population variability. These problems are also present in our approach, but to a markedly reduced degree since the pooled sample representing the population is 3-4 times larger than the target sample for NHANES.

Specifically, the hypothetical population (P) consisted of 31,847 cases, obtained by combining samples from NHANES-I (1971-74, N=11,678), NHANES-II (1976-80, N=10,371), NHANES-III, Phase 1 (1988-91, N=6,874) and HHANES (Mexican Americans only; 1982-84, N=2,944). These counts excluded non-adults (age 0-19) and anyone with missing values on the following ten examination variables, of primary interest in the study:

standing height (ht)
sitting height (sitht)
weight (wt)
systolic blood pressure (sys)
diastolic blood pressure (dia)
total serum cholesterol (tcres)
hemoglobin (hgb)
hematocrit (hct)
iron (iron)
total iron binding capacity (tibc).

The population also included values of the following variables:

age (age)
sex (sex)
race/ethnicity (race, 3 levels)
location (stdrc, 13 levels)
household size (hhsize)
marital status (marr)
education in years (grade)
poverty index (pov)
self report diabetes diagnosed (diab)
self report heart attack diagnosed (htatk)
self report interview height (qht)
self report interview weight (qwt).

Age, sex, race/ethnicity, hhsize and location were fully observed. The other variables had modest numbers of missing values with the exception of qht and qwt, which were not asked in NHANES 1 and hence were missing for all the 11,678 cases from that survey. These and other missing values were imputed using a hot-deck procedure, which did not make strong parametric assumptions about the distribution of the missing values.

2.2 Stratification of the Population. Sample selection from P was achieved by stratified random sampling. Initially units were to be classified into $3\times2\times4\times13$ = 312 strata defined by Race (White/Other, Black, Mexican American) crossed by Sex crossed by Age (20-39, 40-59, 60-74, 75+) crossed by Location [13 categories -- 5 major Stands common to all NHANES (New York, Chicago, Houston, Los Angeles, Philadelphia) and 4 Regions x 2 SMSA categories]. Subsequently this scheme required considerable modification to yield sufficient population cases within each stratum. The final scheme created 48 strata defined by crossing two levels of age (20-59, 60+) with 24 non-overlapping race-geography-location cells.

2.3 Weighting units to create P. Within each stratum h, let N_h denote the estimated U.S. Population Counts at the time of NHANES 97+. These counts were estimated by projecting 1990 Census data to 1997. Let unit i in P have original selection probability π_i in the sample in which it was drawn, including weighting and post-stratification adjustments for nonresponse. Unit i falling in stratum h was assigned weight

$$w_i^* = N_h \frac{(1/\pi_i)}{\sum_{j\in h}(1/\pi_j)},$$

which is the number of members in the hypothetical population that unit i represents. Note that $\sum_{i\in h} w_i^* = N_h$. Also, note that the same formula would hold for a population bootstrapped from NHANES-III, except that the number of units represented by each actual unit, w_i^*, would be approximately 3-4 times as large.

2.4 Drawing Samples from P. The total sample size of the simulated samples was originally planned at about 9000 cases, which is the projected sample size for NHANES 97+. However this number was scaled back to 6000 to avoid excessive bootstrapping, so that the projected number of cases would not exceed one third of the number of actual cases available in each stratum of the pseudo-population. The sample size n_h within stratum h was fixed by allocating the target sample size $n = 6000$ over the strata with proportions based on NHANES-III, Phase 1 data.

If the full population were available, each unit in stratum h would be selected with probability $f_h = n_h / N_h$. Accordingly, unit i in the pseudo-population was sampled m_i times, where m_i was drawn from a binomial distribution with probability f_h and index (that is, count) w_i^* rounded to the nearest integer. This process was repeated on all the units in P, producing a stratified Bernoulli sample with expected sample size n_h in stratum h and $n = \sum_h n_h$ in the entire sample.

When assessing procedures, 1000 samples were drawn from the pseudo-population in this way. The samples have the correct weighting structure of NHANES-III and capture most of the stratification structure, but they do not account for the complexities of clustering of subjects between and within stands. It would require major additional effort to capture this structure, which involves information not currently available for previous NHANES samples.

2.5 Creating Nonresponse. Having drawn a sample, unit, component and item nonresponse were based on patterns of nonresponse from NHANES-III, using the following hot-deck procedure. All interviewed cases in NHANES-III were classified in a contingency table with 576 cells, formed by crossing the 24 sampling strata with sex, age (20-39, 40-59, 60-74, 75+) and household size (1-2, 3-4, 5+). Cells were then collapsed to yield at least five interviewed NHANES-III persons in each cell. Then each sample case was classified in the table and assigned a missing-data pattern from a randomly-drawn NHANES-III case from the same cell. This procedure ignores association of nonresponse with covariates that affect item nonresponse within an examination component, but this approximation is considered acceptable given that item nonresponse is a relatively small aspect of the nonresponse problem.

The resulting missing-data process is ignorable, but creates realistic patterns of unit and component nonresponse, realistic rates of item nonresponse, and incorporates important dependencies on observed variables. If a missing-data method (such as our multiple imputation procedure) works well here, we can be confident that it also works well in the actual population, provided nonresponse is *ignorable*, that is, depends only on values of observed variables. Future work, preferably based on information from follow-up efforts on nonrespondents, should examine the effects of realistic forms of *nonignorable* nonresponse.

2.6 The Multiple Imputation Method. The multiple imputation model involves 17 variables, 10 of primary interest (ht, sitht, wt, sys, dias, tcres, hgb, hct, iron and tobc) and 7 auxiliary variables, namely age in 4 categories (20-39, 40-59, 60-74, 75+)), sex, race, hhsize in 3 categories (1-2, 3-4 and 5+), location with 13 levels, qht and qwt. Twelve of these variables -- the 10 examination variables plus qht and qwt -- are subject to missing values. Age, sex, race, stdrsc and hhsize are treated as categorical in the model. The sample counts are assumed to have a multinomial distribution over the 5-way cross-classification of these variables, a frequency table with 4 x 2 x 3 x 13 x 3 = 936 cells. The model for this table is saturated -- that is, no constraints are placed on the cell probabilities other than that they sum to one. The model specification for this table has no impact on the imputations, because these variables are fully observed.

The 12 continuous variables are assumed to be conditionally multivariate normal given age, sex, race, stdrc and hhsize, with means that vary among the cells of the frequency table, but with a common within-cell covariance matrix. The model for the means is the same for all the continuous variables, and includes the intercept, main effects for age, sex, race, stdrc and hhsize, and the full set of two-way and three-way interactions of age, sex and race. To make normality a more plausible assumption, four of the variables are transformed by taking logarithms of the raw values, and two (iron and tobc) are transformed to the square root scale.

Five sets of multiple imputations of the missing values were generated for this model, by drawing from their posterior predictive distribution, with noninformative (Jeffreys') prior distributions for the location parameters and the within-cell covariance matrix. The computational method was an application of data augmentation, as described in Schafer (1991) or Schafer, Khare and Ezzati-Rice (1993).

Statistical inferences from the multiply-imputed data sets were obtained using standard multiple imputation methods, as discussed in Rubin (1987).

3. PERFORMANCE OF MULTIPLE IMPUTATION.

3.1 Output from the Simulations. For each of 16 summary quantities, the performance of multiple imputation inferences was assessed for estimates for the whole population and for 27 subpopulations defined by Gender, Race/Ethnicity (White, Black, Mexican American), and Age (20-39, 40-59, 60-74, 75+). The summary quantities consisted of the following means and proportions:

mean standing height (ht)
mean weight (wt)
mean sitting height (sitht)
mean systolic blood pressure (sys)
mean diastolic blood pressure (dias)
mean total serum cholesterol (tcres)
mean iron (iron)
mean total iron binding capacity (tibc)
mean hematocrit (hct)
mean hemoglobin (hgb)
proportion hypertensive (hyper)
proportion high cholesterol (hichol)
proportion underweight (uwt)
proportion overweight (owt)
proportion severely overweight (swt)
proportion anemic (anemia)

Operational definitions are as follows:
Body mass index: bmi=$10000 * wt / ht^2$, in kg / m^2; ht in cm.; wt in kg.
Underweight: uwt = 1 if bmi < 20.7 (men), bmi < 19.1 (women), 0 otherwise
Overweight: owt = 1 if bmi > 27.8 (men), bmi >27.3 (women), 0 otherwise
Severely overweight: swt = 1 if bmi >31.1 (men), bmi >32.3 (women), 0 otherwise
Anemia: anemia = 1 if hgb < 13.5 (men), hgb < 12 (women), 0 otherwise
Hypertensive: hyper = 1 if sys ≥ 140 or dias ≥ 90 , 0 otherwise
High cholesterol: hichol = 1 if tcres > 240, 0 otherwise

The following simulation results were recorded for each estimand :
(1) The average means of estimates from the complete sample data before deletion of missing values (BD for before deletion).

(2) The square root of the average estimated variances of estimates from the complete sample data before deletion of missing values.

(3) The average means of estimates from multiple imputation applied to the incomplete data after deletion of missing values (MI).

(4) The square root of the average variance from multiple imputation applied to the incomplete data after deletion of missing values.

(5) The ratio of column (4) to column (2), measuring the increase in estimated standard error from the incomplete data (*relse*).

(6) The number of samples (out of 1000) in which the large sample 95% confidence interval based on the complete cases covered the true population value. Values over 950 imply greater than nominal coverage, under 950 less than nominal coverage.

(7) The number of samples (out of 1000) in which the large sample 95% confidence interval based on multiple imputation of the incomplete cases covered the true population value. Values over 950 imply greater than nominal coverage, under 950 less than nominal coverage.

(8) The average fraction of missing information estimated from the multiple imputes (*fmi*), as discussed in Rubin (1987).

(9) and (10) The median and average degrees of freedom ν in the t approximation for multiple imputation inference, as discussed in Rubin (1987).

For a few estimands involving a very rare binary attribute, the attribute took the value zero for all the cases in at least one of the multiply-imputed data sets, and hence the standard error could not be computed using the usual binomial formula. These problems were excluded in the analysis of the simulation output that follows.

3.2 Summary Results of Simulations. Over all problems, the average coverage of the 95% intervals for MI was 95.5% (sd = 1.8%), compared with an average coverage of 93.9% (sd = 1.4%) for BD. Since the complete-data intervals have coverages slightly below the nominal level, the average coverage of MI is actually better than the coverages from BD. On the other hand, the coverages of the multiple imputation intervals are slightly more variable than the coverages of the complete data intervals. Table 1A summarizes the BD and MI coverages by variable, indicating that MI provides consistent coverage for all the variables considered.

The fraction of missing information estimated from multiple imputation (*fmi*) averaged 17.6% over all estimands, with a standard deviation of 8.3%. To assess how well this measure reflected the true increase in standard error from the missing data, a variable $comp = relse / \sqrt{1 - fmi}$ was computed for each estimand. This variable takes the value one when *fmi* is a perfect measure of the loss of information. Over all problems, *comp* averaged 1.01, with a standard deviation of .03, suggesting that *fmi* is indeed a useful measure of information loss.

To assess whether the coverage of MI was systematically related to the fraction of missing information, problems were grouped into 5 categories based on values of relse (namely, less than 1.05, 1.05 to 1.1, 1.1 to 1.2, 1.2 to 1.3, and greater than or equal to 1.3). Table 1B summarizes the coverages of MI and BD in these five classes. The coverages of BD show no systematic pattern. The coverages of MI increase slightly with relse,

Table 1. Mean (sd) of Coverages of Intervals from Complete Case Before Deletion (BD) and Multiple Imputation (MI), Classified by (A) Variables; (B) Ratio of Standard Errors.

(A) By Variable

	HT	WT	SITHT	SYS
CC	944 (14)	941 (17)	944 (14)	943 (18)
MI	951 (15)	951 (17)	950 (16)	950 (22)
	DIAS	TCRES	IRON	TIBC
CC	947 (15)	942 (13)	943 (12)	944 (15)
MI	954 (18)	954 (16)	952 (16)	955 (16)
	HCT	HGB	HYPER	HICHOL
CC	941 (18)	942 (15)	943 (15)	943 (11)
MI	951 (21)	951 (20)	961 (13)	961 (13)
	UWT	OWT	SWT	ANEMIA
CC	940 (18)	945 (9)	938 (12)	940 (19)
MI	962 (24)	960 (10)	959 (18)	958 (25)

(B) By Ratio of Standard Errors of MI to CC.

	RATSE				
	<1.05	1.05-1.10	1.10-1.20	1.20-1.30	>1.30
CC	940 (17)	943 (12)	944 (15)	943 (16)	947 (17)
MI	948 (21)	956 (17)	957 (17)	955 (17)	959 (17)

reflecting the conservative nature of the MI adjustment. The degree of conservatism is slight, however, with coverages increasing from 94.8% to 95.9% over the five classes.

The median degrees of freedom from the multiple imputation inferences were consistently large (data not shown), suggesting that (at least for these estimands) inference based on a normal approximation will be acceptable.

3.3 Conclusion and Future Work. The simulation results described here suggest that our multiple imputation method provide reliable inferences for a variety of means and proportions, for the whole sample and for subclasses defined by Gender, Race/Ethnicity and Age. The range of estimands considered in this report is limited, and does not involve more complex statistics such as regression coefficients and percentiles. Simulations were also conducted on percentiles, but are not presented here because the complete-data methods were non-standard and complete-data coverage was not always satisfactory. Further work will analyze these results in more detail, and include more research on estimands such as regression coefficients that may be more vulnerable to misspecification of the imputation model.

ACKNOWLEDGMENTS

The work of Little, Rubin and Schafer on this project was partially supported by professional services contracts from the National Center for Health Statistics, Centers for Disease Control and Prevention, Department of Health and Human Services. Authors are listed alphabetically.

REFERENCES

Ezzati, T. and Khare, M. (1992), "Nonresponse adjustments in a National Health Survey", *Proceedings of the Survey Research Methods Section, American Statistical Association 1992*, 339-344.

Ezzati-Rice, T.M., Khare, M., Rubin, D.B., Little, R.J.A. and Schafer, J.L. (1993), "A Comparison of Imputation Techniques in the Third National Health and Nutrition Examination Survey", *Proceedings of the Survey Research Methods Section, American Statistical Association 1993*, in press.

Fahimi, M. and Judkins, D. (1993), "Serial Imputation of NHANES III with Mixed Regression and Hot-Deck Techniques", *Proceedings of the Survey Research Methods Section, American Statistical Association 1993*, in press.

Khare, M., Little, R.J.A., Rubin, D.B., and Schafer, J.L. (1993), "Multiple Imputation of NHANES III", *Proceedings of the Survey Research Methods Section, American Statistical Association 1993*, in press.

Little, R.J.A. and Rubin, D.B. (1992), "Assessment of Trial Imputations for NHANES III", Final Report, Datametrics Research, Inc., Waban: Massachusetts.

Little, R.J.A. and Schluchter, M.D. (1985), "Maximum likelihood estimation for mixed continuous and categorical data with missing values," *Biometrika,* 72, 497-512.

Rubin, D.B. (1987), *Multiple Imputation for Nonresponse in Surveys*, New York: John Wiley and Sons.

Schafer, J. (1991), *Algorithms for Multiple Imputation and Posterior Simulation from Incomplete Multivariate Data with Ignorable Nonresponse,* Ph.D. Thesis, Department of Statistics, Harvard University.

Schafer, J.L., Khare, M. and Ezzati-Rice, J.M. (1993), "Multiple imputation of missing data in NHANES III," *Proceedings of the 1993 Annual Research Conference*, U.S. Bureau of the Census, Washington, D.C, 459-487.